

Gaussian process regression with functional covariates and multivariate response

Wang, B. , Chen, T. and Xu, A.

Post-print deposited in Coventry University repository

Original citation:

Wang, B. , Chen, T. and Xu, A. (2017) Gaussian process regression with functional covariates and multivariate response. Chemometrics and Intelligent Laboratory Systems, volume 163 : 1-6. DOI: 10.1016/j.chemolab.2017.02.001

<http://dx.doi.org/10.1016/j.chemolab.2017.02.001>

Elsevier

Creative Commons Attribution Non-Commercial No Derivatives License

Gaussian process regression with functional covariates and multivariate response

Bo Wang^{a,*}, Tao Chen^b, Aiping Xu^c

^a*Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK*

^b*Department of Chemical and Process Engineering, University of Surrey, Guildford GU2 7XH, UK*

^c*Sigma (Maths and Stats Support), Coventry University, Coventry CV1 5DD, UK*

Abstract

Gaussian process regression (GPR) has been shown to be a powerful and effective non-parametric method for regression, classification and interpolation, due to many of its desirable properties. However, most GPR models consider univariate or multivariate covariates only. In this paper we extend the GPR models to cases where the covariates include both functional and multivariate variables and the response is multidimensional. The model naturally incorporates two different types of covariates: multivariate and functional, and the principal component analysis is used to de-correlate the multivariate response which avoids the widely recognised difficulty in the multi-output GPR models of formulating covariance functions which have to describe the correlations not only between data points but also between responses. The usefulness of the proposed method is demonstrated through a simulated example and two real data sets in chemometrics.

Keywords: Gaussian process regression, functional data analysis, functional covariates, multivariate response, semi-metrics

1. Introduction

Gaussian process regression (GPR), first proposed by [1] and further developed in machine learning community [2, 3], has received increasing interests in recent years,

*Corresponding author. Tel.: +44 116 252 2162, Fax: +44 116 252 3915.

Email addresses: bo.wang@le.ac.uk (Bo Wang), t.chen@surrey.ac.uk (Tao Chen), aa9778@coventry.ac.uk (Aiping Xu)

due to many of its desirable properties, such as the existence of explicit forms, the ease of obtaining and expressing uncertainty in predictions, the ability to capture a wide variety of behaviour through covariance functions, and a natural Bayesian interpretation. It has been shown to be a powerful and effective nonparametric method for regression, classification and interpolation by various empirical studies in a wide range of fields. See for example the seminal book [3] and the references therein for details. In chemometrics and related areas, GPR has been applied to a range of problems, such as calibration of spectroscopic analysers [4, 5], response surface modelling [6], system identification [7], ensemble learning [5, 8], prediction of transmembrane pressure [9], and prediction of percutaneous absorption [10, 11], among others.

However, in the literature the covariates involved in Gaussian process methods are mostly univariate or multivariate, such as time, spatial location, or multivariate measurements in chemical processes. In this paper we extend the GPR model to the cases where the covariates contain infinite dimensional variables or functional data. Functional data concerns data which are collected as curves, surfaces or measurements varying over a continuum. With the fast advances of technologies which allow to record data with great accuracy and at high frequency, functional data have become more and more prevalent in an increasing number of areas, such as biology, engineering, medical science, meteorology, psychology, statistics, among others. Although multivariate statistical methods may be applied to some of such data (which is not always the case, for example, when repeated measurements are sampled at different time points), there are several advantages treating such data as functional. For instance, functional data treatment enables us to extract additional information contained in the data, such as the order information of observations and the derivatives (rates of changes, slope, curvature, etc); see Levitin et al [12] for more discussion on this aspect.

The statistical methods for analysing functional data, termed as *functional data analysis* (FDA), was pioneered by [13] and [14] and have experienced fast development in recent years. FDA has also been applied in many chemometrical problems. One of the classical examples in FDA is the spectrometric data, which concerns the absorbance

measured at a grid of different wavelengths for a sample of finely chopped meat and has been studied by [15] and many others. Other applications of FDA in chemometrics include [16], [17], and [18].

The purpose of this paper is two-fold. Firstly, we extend the GPR models to the situation where the covariates can be functional as well as multivariate. Secondly, we deal with the regression problem where the response variable is multidimensional. Multi-response GPR models with multivariate covariates have been studied by several authors and various methods have been proposed. For example, [5, 6] treat each response variable as a Gaussian process and multiple responses are modelled independently; [19] treats Gaussian processes as the outputs of stable linear filters; [20] proposes a direct formulation of the covariance function for multi-response GPR, based on the idea that its covariance function is assumed to be the “nominal” uni-output covariance multiplied by the covariances between different outputs. In this paper a new method is proposed to deal with correlated multivariate response variables with multivariate and functional covariates. We use the principal component analysis (PCA) to de-correlate the multivariate response which then enables us to model each principal component independently by a Gaussian process regression. In the GPR models, the closeness between the multivariate covariates is calculated by the usual Euclidean distance, while that between the functional covariates is measured by semi-metrics which will be explained in more details later.

The rest of the paper is organised as follows. Section 2 briefly reviews the Gaussian process regression, followed by the presentation of the proposed Gaussian process methods for functional regression with multivariate response. The usefulness of the proposed method will be illustrated through numerical examples in Section 3. And the paper is concluded by Section 4.

2. Methodology

2.1. Gaussian process regression

Let $y \in \mathbb{R}$ be a response variable and $x \in \mathbb{R}^p$ the covariate variable. Consider the following nonlinear regression

$$y = f(x) + \varepsilon$$

where the function $f(\cdot) : \mathbb{R}^p \mapsto \mathbb{R}$ is unknown. By Gaussian process regression method $f(\cdot)$ is assumed to follow a Gaussian process with mean function $\mu(\cdot)$ and covariance function $k(\cdot, \cdot)$. Therefore, if n pairs of data points $(x_1, y_1), \dots, (x_n, y_n)$ are observed, we have

$$y_i = f(x_i) + \varepsilon_i,$$

where $\{\varepsilon_i\}_{i=1, \dots, n}$ are independent and identically distributed normal random noises with mean 0 and variance σ^2 . Hence $(y_1, \dots, y_n)^T$ follows an n -variate normal distribution

$$(y_1, \dots, y_n)^T \sim N(\boldsymbol{\mu}, K),$$

where $\boldsymbol{\mu} = (\mu(x_1), \dots, \mu(x_n))^T$ is the mean vector and K is the $n \times n$ covariance matrix whose (i, j) -th element $K_{ij} = k(x_i, x_j) + \sigma^2 \delta_{ij}$. Here $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Suppose that x^* is a test point and y^* the corresponding response value. Then by the Gaussian process assumption the joint distribution of $(y_1, \dots, y_n, y^*)^T$ is an $(n+1)$ -variate normal distribution with mean $(\mu(x_1), \dots, \mu(x_n), \mu(x^*))^T$ and covariance matrix

$$\begin{bmatrix} K & K^* \\ K^{*T} & k(x^*, x^*) + \sigma^2 \end{bmatrix}$$

where $K^* = (k(x^*, x_1), \dots, k(x^*, x_n))^T$. Therefore the conditional distribution of y^* , given $\mathbf{y} = (y_1, \dots, y_n)^T$, is $N(\hat{y}^*, \hat{\sigma}^{*2})$ where

$$\begin{aligned} \hat{y}^* &= \mu(x^*) + K^{*T} K^{-1}(\mathbf{y} - \boldsymbol{\mu}), \\ \hat{\sigma}^{*2} &= k(x^*, x^*) + \sigma^2 - K^{*T} K^{-1} K^*. \end{aligned}$$

In Gaussian process regression, the covariance function $k(\cdot, \cdot)$ reflects our presumptions about the unknown function so plays an important role. A number of different

covariance functions have been proposed and discussed in the literature; see for example [3] and [21]. In this paper we will adopt the most commonly used covariance function - the squared exponential covariance function:

$$k(x_i, x_j) = v \exp \left(-\frac{1}{2} \sum_{d=1}^p w_d (x_{id} - x_{jd})^2 \right), \quad (1)$$

and assume the mean function $\mu(x)$ to be 0.

The hyper-parameters $\{v, w_1, \dots, w_p\}$ in (1) and the noise variance σ^2 can be estimated by the maximum likelihood method. The log-likelihood of the training data is given by

$$L(v, w_1, \dots, w_p, \sigma^2) = -\frac{1}{2} \log \det K - \frac{1}{2} \mathbf{y}^T K^{-1} \mathbf{y} - \frac{n}{2} \log 2\pi.$$

The derivative of the log-likelihood with respect to each parameter (denoted by a generic notation θ) is given by:

$$\frac{\partial L}{\partial \theta} = -\frac{1}{2} \text{tr} \left(K^{-1} \frac{\partial K}{\partial \theta} \right) + \frac{1}{2} \mathbf{y}^T K^{-1} \frac{\partial K}{\partial \theta} K^{-1} \mathbf{y}.$$

2.2. Gaussian process regression with functional covariates and multivariate response

Now suppose that $Y = (y_1, \dots, y_m)^T$ is a multivariate response in \mathbb{R}^m , $X(t)$ a q -dimensional functional covariate, and z a p -dimensional multivariate covariate. Consider the problem of nonlinear regression

$$Y = \mathbf{f}(X(t), z) + \boldsymbol{\varepsilon}$$

where $\mathbf{f} = (f_1, \dots, f_m)^T$ is an unknown m -dimensional function, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_m)^T \sim N(\mathbf{0}, \Sigma)$ with Σ being an $m \times m$ covariance matrix.

Given n observations $(X_1, z_1, Y_1), \dots, (X_n, z_n, Y_n)$, let $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ be the sample mean vector and the sample covariance matrix of $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, respectively, and have the eigenvalue-(normalised) eigenvector pairs $(\lambda_1, \mathbf{e}_1), \dots, (\lambda_m, \mathbf{e}_m)$ where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$. Then by the principal component analysis the principal scores are given by

$$\boldsymbol{\Psi} = (\mathbf{Y} - \mathbf{U})\mathbf{V},$$

where $\mathbf{V} = (\mathbf{e}_1, \dots, \mathbf{e}_m)$, and $\mathbf{U} = (\hat{\boldsymbol{\mu}}, \dots, \hat{\boldsymbol{\mu}})^T$ is an $n \times m$ matrix.

Denote by $\boldsymbol{\psi}_l = (\psi_{1l}, \dots, \psi_{nl})^T$ the l th column of $\boldsymbol{\Psi}$, then $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_m$ are samples of m uncorrelated random variables. Therefore the regression function $\mathbf{f}(\cdot, \cdot)$ can be represented by the relationships between ψ_{il} and $X_i(t)$ and z_i , that is, for $l = 1, \dots, m$ and $i = 1, \dots, n$,

$$\psi_{il} = r_l(X_i(t), z_i) + e_{il}, \quad (2)$$

where $e_{il} \sim N(0, \sigma_l^2)$. By Gaussian process methods, we assume $r_l(\cdot, \cdot)$ to follow a Gaussian process with mean function 0 and covariance function $k^{(l)}(\cdot, \cdot, \cdot, \cdot)$, defined as

$$k^{(l)}(X_i, X_j, z_i, z_j) = v^{(l)} \exp \left(-\frac{1}{2} \sum_{d=1}^p w_d^{(l)} (z_{id} - z_{jd})^2 - \frac{1}{2} \sum_{d=1}^q \eta_d^{(l)} \|X_{id} - X_{jd}\|_d^2 \right), \quad (3)$$

where $\|\cdot\|_d$ denotes the semi-metric defined for the d th component of the functional covariate. Semi-metrics as a closeness measure for functional data were introduced in [15], and a number of different semi-metrics have been proposed in the literature. In this paper two semi-metrics, which are based on functional principal component analysis (FPCA) and on derivatives, are employed in the numerical examples and are briefly presented below for completeness.

Let $\mathcal{X}_1, \dots, \mathcal{X}_n$ be a sample of curves which are identically distributed as the functional random variable $\mathcal{X} = \{\mathcal{X}(t); t \in \mathcal{T}\}$.

Semi-metric based on FPCA is defined as

$$\|\mathcal{X}_i - \mathcal{X}_j\|^{FPCA, q} = \sqrt{\sum_{k=1}^q \left(\int [\mathcal{X}_i(t) - \mathcal{X}_j(t)] v_k(t) dt \right)^2},$$

where v_1, \dots, v_q are the orthonormal eigenfunctions of the covariance operator $\Gamma_{\mathcal{X}}(s, t) = E(\mathcal{X}(s)\mathcal{X}(t))$ associated with the largest q eigenvalues.

Semi-metric based on derivatives is defined as

$$\|\mathcal{X}_i - \mathcal{X}_j\|^{deriv, q} = \sqrt{\int \left(\mathcal{X}_i^{(q)}(t) - \mathcal{X}_j^{(q)}(t) \right)^2 dt},$$

where $\mathcal{X}^{(q)}$ denotes the q th derivative of \mathcal{X} with respect to t , which is computed using the B-spline approximation of the curves in practice.

The first semi-metric is suitable for rough functional data, while the derivatives-type semi-metric is adapted to smooth functions. We refer to [15] for the practical implementation of the semi-metrics. It should be noted that the semi-metrics in (3) can be chosen differently for different components of the functional covariate as appropriate.

The model training and prediction for (2) and (3) can be performed in the same way as presented in the previous subsection. Thus given a test input $(X^*(t), z^*)$, the predictive means and variances of the scores can then be obtained and denoted by $\hat{\psi}_l^*$ and $\hat{\sigma}_l^{*2}$ for $l = 1, \dots, m$. Consequently the predictive mean and variance of the m -dimensional response Y^* are given by

$$E(Y^*) = \hat{\boldsymbol{\mu}} + \mathbf{V}\boldsymbol{\psi}^*, \quad (4)$$

$$Var(Y^*) = Var(\hat{\boldsymbol{\mu}}) + \mathbf{V}\Sigma^*\mathbf{V}^T + \hat{\Sigma}, \quad (5)$$

where $\boldsymbol{\psi}^* = (\hat{\psi}_1^*, \dots, \hat{\psi}_m^*)^T$, $\Sigma^* = \text{diag}(\hat{\sigma}_1^{*2}, \dots, \hat{\sigma}_m^{*2})$, and $Var(\hat{\boldsymbol{\mu}}) = \hat{\Sigma}/n$.

3. Numerical examples

In this section we demonstrate the usefulness of the proposed method using some numerical examples, including simulated data and real data.

3.1. Simulated example

We first consider a simulated example, which is adapted from [22]. The data are generated as follows. Let X_1, \dots, X_n be $n = 150$ samples of a functional covariate such that

$$X_i(t_j) = a_i \cos(2t_j) + b_i \sin(4t_j) + c_i(t_j^2 - \pi t_j + \frac{2}{9}\pi^2) + \epsilon_{ij},$$

where a_i , b_i and c_i ($i = 1, \dots, n$) are all independent real random variables uniformly distributed in $[0, 1]$, $\epsilon_{ij} \sim N(0, 0.01^2)$ and $0 = t_1 < t_2 < \dots < t_{100} = \pi$ are equally spaced points.

The regression functions and the response variables are defined as, for $i = 1, \dots, n$,

$$f_1(X_i) = \frac{1}{100} \int_0^\pi X_i^2(t) dt, \quad y_{1i} = f_1(X_i) + \varepsilon_{1i},$$

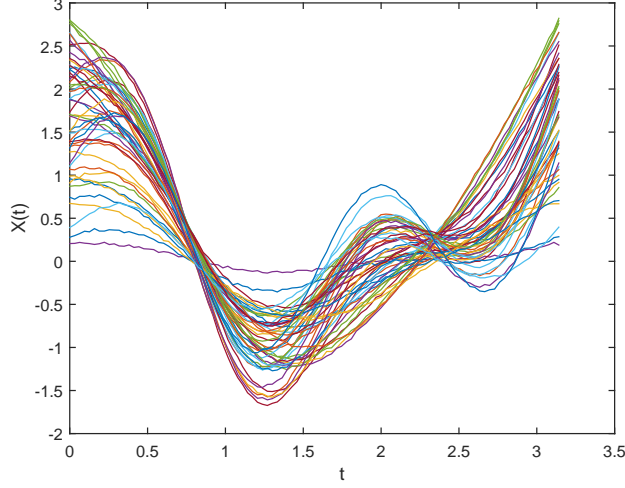


Figure 1: Sample covariate curves.

$$f_2(X_i) = \frac{1}{100} \int_0^\pi t \cos(t) X_i^2(t) dt, \quad y_{2i} = f_2(X_i) + \varepsilon_{2i},$$

with

$$\boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N(\mathbf{0}, \Sigma), \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

We use $\sigma_1 = 0.1$, $\sigma_2 = 0.12$ and $\rho = 0.9$ or 0.2 to represent strong or weak correlation between the two response variables, respectively. An example of fifty covariate curves is shown in Figure 1.

We split the sample into two sets: the training sample $(X_i, Y_i)_{i=1, \dots, 100}$ and the test sample $(X_i, Y_i)_{i=101, \dots, 150}$. Since the covariate curves are not smooth, the semi-metric based on FPCA is adopted in our model. The root mean square error (RMSE) between the predicted and the true regression values for the fifty test samples is used as a measure of prediction accuracy. The performance of the proposed method (Multi-GP) is compared with that of the independent models (Ind-GP) where the two response variables are modelled independently and without considering their correlation. The above process is repeated 20 times and the average RMSEs are reported in Table 1. It can be observed that the proposed model (Multi-GP) considerably improves the prediction accuracy compared with the method of modelling each response independently

Table 1: The average RMSEs for the simulated data

	Response 1 (y_1)		Response 2 (y_2)	
	Multi-GP	Ind-GP	Multi-GP	Ind-GP
$\rho = 0.9$	0.0292	0.0350	0.0346	0.0385
$\rho = 0.2$	0.0336	0.0350	0.0336	0.0338

(Ind-GP), when the correlation between the two response variables is strong. On the other hand, when the correlation is weak the improvement by Multi-GP is marginal in comparison with the independent method.

3.2. Application to Real data

We now apply the proposed method to two real data sets in chemometrics.

Tecator data. The data are recorded on a Tecator Infratec Food and Feed Analyzer working in the wavelength range 850 - 1050nm by the Near Infrared Transmission (NIT) principle. Each sample contains one of 215 finely chopped pure meat with different moisture, fat and protein contents. For each meat sample the data consists of a 100 channel spectrum of absorbances and the contents of water, fat and protein. The absorbance is $-\log_{10}$ of the transmittance measured by the spectrometer. The spectrometric curves are shown in Figure 2. The sample correlations between fat and water and between fat and protein are -0.9881 and -0.8609, respectively, and that between water and protein is 0.8145. The task is to predict the three contents from the spectrometric curves.

As the spectrometric curves are smooth, the semi-metric based on derivative of order two is adopted in our model. To compare the performance, leave-one-out cross validation is performed, that is, each of the 215 samples is left as test data whilst the remaining data are used for model training. The predicted values are then compared with the measured and the root mean square errors by the proposed Multi-GP with functional covariate and the independent models with functional covariate (Ind-GP)

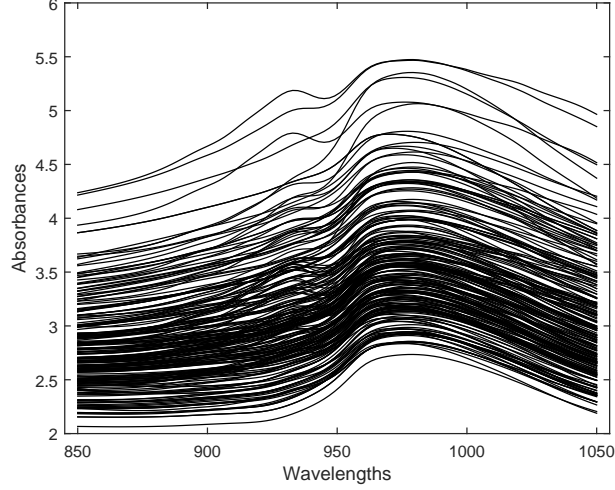


Figure 2: A graphical display of spectrometric curves.

Table 2: The RMSEs for the Tecator data

	Fat	Water	Protein
Multi-GP	1.429	1.548	0.942
Ind-GP	1.450	1.816	0.923
GP-MV	1.724	1.516	1.222

are reported in Table 2. The predictions by both methods are illustrated in Figure 3.

Although many functional data can not be analysed by multivariate statistical methods as pointed out earlier, it is, nevertheless, possible to treat the spectrometric curves as multivariate in this example. Therefore, for demonstration we also compare the performance of GPR with functional covariate with that of the conventional GPR with multivariate covariates (GP-MV) where each response is independently modelled by a GPR with the covariates being the absorbances at each wavelength. Since the number of wavelengths is large (100 in total), the usual strategy of PCA is adopted to reduce the number of covariates and five principal components, which account for 99.996% of the total variation, is used in the GPR models as the covariates. The

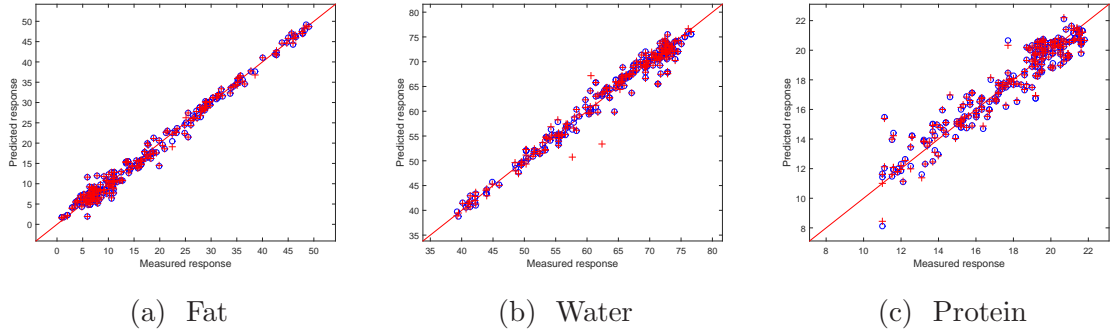


Figure 3: Prediction by leave-one-out cross validation for Tecator data. 'o': by Multi-GP; '+': by Ind-GP.

RMSEs of the predictions by this model are also shown in Table 2 as GP-MV.

It can be seen that, in comparison with the independent model, the Multi-GP method significantly improves the prediction accuracy for water whilst there is marginal improvement for fat. The prediction for protein by Multi-GP is slightly worse than Ind-GP, but in overall terms the former outperforms the independent model. Although GP-MV provides the best prediction for water, its performance on fat and protein is notably worse than the other two methods.

Soil data. This data set was originally studied by [23]. The soil samples were obtained at a long-term field experiment in Abisko, northern Sweden. The samples are from 36 plots, with three subsamples from each plot, giving a total of 108 samples. The wavelength range of 400 - 2500nm (visible and near infrared spectrum) was scanned at 2nm intervals with an NIR spectrophotometer and fluorescence excitation-emission matrices (EEMs) were recorded with a spectrofluorometer. Two reference values, soil organic matter (SOM) was measured as loss on ignition at 550 °C, and ergosterol concentration (EC) was determined through HPLC. The sample correlation between the two reference values is 0.7108.

The semi-metric based on derivative of order two is adopted in our model since the covariate curves are smooth, and the same leave-one-out cross validation experiment

Table 3: The RMSEs for the Soil data

	SOM	EC
Multi-GP	2.380	36.329
Ind-GP	2.432	37.098
GP-MV	2.942	36.743

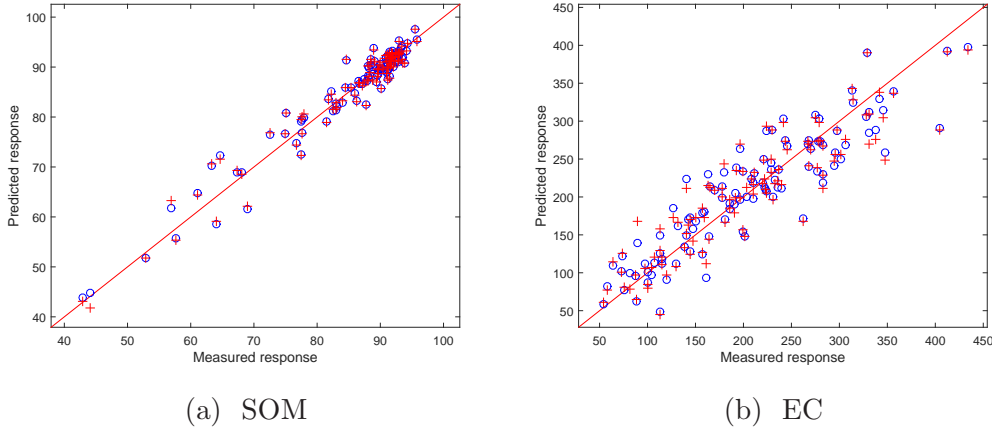


Figure 4: Prediction by leave-one-out cross validation for Soil data. 'o': by Multi-GP; '+': by Ind-GP.

as in the previous example is performed. The root mean square errors by Multi-GP, Ind-GP and GP-MV (with five principal components accounting for 99.789% of the total variation) are reported in Table 3, and the predictions by the first two methods are shown in Figure 4.

It can be seen that the proposed Multi-GP model improves the accuracy of prediction for both SOM and EC in comparison with the Ind-GP and GP-MV.

The above numerical examples also manifest another advantage of functional data over multivariate, in addition to what has been discussed in the introduction, that is, treating some data as functional rather than multivariate in GPR modelling significantly reduces the model complexity. For instance, only one parameter is needed for

the spectrometric curves in the soil data, instead of 1050 if each wavelength is treated as a variable or 5-10 if PCA is used to reduce the dimensionality.

4. Conclusion

In this paper we extend the Gaussian process regression models to the situation where the covariates include functional variable as well as multivariate variable and with multidimensional response. The closeness between the functional covariates is measured by the semi-metrics as discussed in [15]. Principal component analysis is performed for the multivariate response and the uncorrelated principal scores are then modelled using Gaussian process regressions. The usefulness of the proposed method over the independent model and the conventional GPR models is demonstrated through some numerical examples.

The proposed Gaussian process model naturally incorporates two different types of covariates: multivariate and functional. And the PCA technique to deal with multivariate response avoids the widely recognised difficulty in GPR for multiple outputs, that is, to formulate a covariance function that describes not only the correlation between data points, but also the correlation between responses. In this paper we assume that the dimension of the response is low so that all the principal components are used in the subsequent regression models. However, if the response is high dimensional and it is intractable to build a GPR model for each of the principal components, we can choose to model fewer number of principal components, for example choosing the minimum number of components which explain at least 95% of the total variation, or, determining this number by cross validation.

Conflict of interest

The authors confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgement

The authors thank the reviewers for their constructive suggestions and helpful comments.

References

- [1] A. O’Hagan, Curve fitting and optimal design for prediction, *J. Roy. Statist. Soc. Ser. B* 40 (1) (1978) 1–42.
- [2] D. J. C. MacKay, Introduction to Gaussian processes, Tech. rep., Cambridge University (1999).
- [3] C. E. Rasmussen, C. K. I. Williams, *Gaussian Processes For Machine Learning*, The MIT Press, 2006.
- [4] T. Chen, J. Morris, E. Martin, Gaussian process regression for multivariate spectroscopic calibration, *Chemometrics and Intelligent Laboratory Systems* 87 (2007) 59–67.
- [5] K. Wang, T. Chen, R. Lau, Bagging for robust non-linear multivariate calibration of spectroscopy, *Chemometrics and Intelligent Laboratory Systems* 105 (2011) 1–6.
- [6] J. Yuan, K. Wang, T. Yu, M. Fang, Reliable multi-objective optimization of high-speed WEDM process based on Gaussian process regression, *International Journal of Machine Tools & Manufacture* 48 (2008) 47–60.
- [7] L. L. T. Chan, Y. Liu, J. Chen, Online system identification with selective recursive Gaussian process models, *Industrial & Engineering Chemistry Research* 52 (2013) 18276–18286.
- [8] Y. Liu, Z. Gao, Real-time property prediction for an industrial rubber-mixing process with probabilistic ensemble Gaussian process regression models, *Journal of Applied Polymer Science* 132 (6) (2015) 1905–1913.

- [9] Y. Liu, C.-P. Chou, J. Chen, J.-Y. Lai, Active learning assisted strategy of constructing hybrid models in repetitive operations of membrane filtration processes: Using case of mixture of bentonite clay and sodium alginate, *Journal of Membrane Science* 515 (2016) 245–257.
- [10] G. P. Moss, Y. Sun, M. Prapopoulou, N. Davey, R. Adams, W. J. Pugh, M. B. Brown, The application of Gaussian processes in the prediction of percutaneous absorption, *Journal of Pharmacy and Pharmacology* 61 (9) (2009) 1147–1153.
- [11] L. T. Lam, Y. Sun, N. Davey, R. Adams, M. Prapopoulou, M. B. Brown, G. P. Moss, The application of feature selection to the development of Gaussian process models for percutaneous absorption, *Journal of Pharmacy and Pharmacology* 62 (6) (2010) 738–749.
- [12] D. J. Levitin, R. L. Nuzzo, B. W. Vines, J. Ramsay, Introduction to functional data analysis, *Canadian Psychology/Psychologie Canadienne* 48 (3) (2007) 135–155.
- [13] J. O. Ramsay, C. J. Dalzell, Some tools for functional data analysis, *Journal of the Royal Statistical Society. Series B (Methodological)* (1991) 539–572.
- [14] J. O. Ramsay, B. W. Silverman, *Functional Data Analysis*, Springer, New York, 2005.
- [15] F. Ferraty, P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, New York, 2006.
- [16] W. Saeys, B. De Ketelaere, P. Darius, Potential applications of functional data analysis in chemometrics, *Journal of Chemometrics* 22 (5) (2008) 335–344.
- [17] A. Aguilera, M. Escabias, M. Valderrama, M. Aguilera-Morillo, Functional analysis of chemometric data, *Open Journal of Statistics* 2013 (3) (2013) 334–343.

- [18] M. Muller, Functional data analysis applied in chemometrics: With focus on NMR nutri-metabolomics, Ph.D. thesis, University of Copenhagen (2015).
- [19] P. Boyle, M. Frean, Dependent Gaussian processes, in: Advances in Neural Information Processing Systems, 2004, pp. 217–224.
- [20] B. Wang, T. Chen, Gaussian process regression with multiple response variables, Chemometrics and Intelligent Laboratory Systems 142 (2015) 159–165.
- [21] J. Q. Shi, T. Choi, Gaussian process regression analysis for functional data, CRC Press, Boca Raton, FL, 2011.
- [22] F. Ferraty, I. Van Keilegom, P. Vieu, On the validity of the bootstrap in non-parametric functional regression, Scandinavian Journal of Statistics 37 (2) (2010) 286–306.
- [23] R. Rinnan, A. Rinnan, Application of near infrared reflectance (NIR) and fluorescence spectroscopy to analysis of microbiological and chemical properties of arctic soil, Soil Biology and Biochemistry 39 (7) (2007) 1664–1673.

List of Figures

1	Sample covariate curves.	8
2	A graphical display of spectrometric curves.	10
3	Prediction by leave-one-out cross validation for Tecator data. 'o': by Multi-GP; '+' : by Ind-GP.	11
4	Prediction by leave-one-out cross validation for Soil data. 'o': by Multi- GP; '+' : by Ind-GP.	12

List of Tables

1	The average RMSEs for the simulated data	9
2	The RMSEs for the Tecator data	10
3	The RMSEs for the Soil data	12